Shared Linguistic Resources for the Meeting Domain

Meghan Lammie Glenn, Stephanie Strassel

Linguistic Data Consortium
3600 Market Street, Suite 810
Philadelphia, PA 19104
{mlglenn, strassel}@ldc.upenn.edu

Abstract. This paper describes efforts by the University of Pennsylvania's Linguistic Data Consortium to create and distribute shared linguistic resources – including data, annotations, tools and infrastructure – to support the Spring 2007 (RT-07) Rich Transcription Meeting Recognition Evaluation. In addition to making available large volumes of training data to research participants, LDC produced reference transcripts for the NIST Phase II Corpus and RT-07 conference room evaluation set, which represent a variety of subjects, scenarios and recording conditions. For the 18-hour NIST Phase II Corpus, LDC created quick transcripts which include automatic segmentation and minimal markup. The 3-hour evaluation corpus required the creation of careful verbatim reference transcripts including manual segmentation and rich markup. The 2007 effort marked the second year of using the XTrans annotation tool kit in the meeting domain. We describe the process of creating transcripts for the RT-07 evaluation, and describe the advantages of utilizing XTrans for each phase of transcription and its positive impact on quality control and real-time transcription rates. This paper also describes the structure and results of a pilot consistency study that we conducted on the 3-hour test set. Finally, we present plans for further improvements to infrastructure and transcription methods.

Keywords: linguistic resources, transcription, annotation tools, XTrans, Annotation Tool Graph Kit (AGTK)

1 Introduction

Linguistic Data Consortium was established in 1992 at the University of Pennsylvania to support language-related education, research and technology development by creating and sharing linguistic resources, including data, tools and standards. Human language technology development in particular requires large volumes of annotated data for building language models, training systems and evaluating system performance against a human-generated gold standard. LDC has directly supported the National Institute of Standards and Technology's (NIST) Rich Transcription evaluation series by providing training and evaluation data and related infrastructure.

For the Spring 2007 (RT-07) Rich Transcription Meeting Recognition Evaluation, LDC provided large quantities of training data from a variety of domains to program participants. LDC produced 18 hours of new quick transcripts for the NIST Phase II conference room corpus. In addition to that, LDC produced 3 hours of careful reference transcripts of evaluation data to support automatic speech-to-text transcription, diarization, and speaker segmentation and localization in the meeting domain. The RT-07 conference room sets were created by using XTrans, the specialized speech annotation tool that was developed to respond to unique challenges presented by transcription. XTrans supports rapid, high-quality creation of rich transcripts, in the meeting domain and in a wide variety of other genres. It also provides built-in quality control mechanisms that facilitate consistency and improve real-time transcription rates, thereby opening avenues for further experimentation in the reference transcript creation process. This paper also describes a pilot study conducted to begin to understand inter-transcriber consistency. The results show that there are

2 Data

2.1 Training Data

To enhance availability of high-quality training data for RT-07, LDC coordinated with NIST to distribute eight corpora that are part of the LDC catalog for use as training data by evaluation participants. The data included five corpora in the meeting domain and two large corpora of transcribed conversational telephone speech (CTS) as well as one corpus of transcribed broadcast news (BN). All data was shipped directly to registered evaluation participants upon request, after sites had signed a user agreement specifying research use of the data. The distributed training data is summarized in the table below.

Title	Speech	Transcripts	Volume	Domain
Fisher English Part 1	LDC2004S13	LDC2004T19	750+ hours	CTS
Fisher English Part 2	LDC2005S13	LDC2005T19	750+ hours	CTS
ICSI Meeting Corpus	LDC2004S02	LDC2004T04	72 hours	Meeting
ISL Meeting Corpus	LDC2004S05	LDC2004T10	10 hours	Meeting
NIST Meeting Pilot Corpus	LDC2004S09	LDC2004T13	13 hours	Meeting
RT-04S Dev-Eval Meeting Room Data	LDC2005S09	LDC2005S09	14.5 hours	Meeting
RT-06 Spring Meeting Speech Evaluation Data		LDC2006E16	3 hours	Meeting
TDT4 Multilingual Broadcast News Corpus	LDC2005S11	LDC2005T16	300+ hours	BN

Table 1. RT-07S Training Data Distributed through NIST by LDC

2.2 NIST Phase II Data

LDC transcribed 18 hours of meeting recordings for the NIST Phase II Corpus, using the Quick Transcription (QTR) methodology. The corpus is comprised of 17 files, ranging from 40 minutes to nearly 2 hours in duration. There are between 3 and 6 speakers per session, including native and non-native speakers, and 2 "ambient" speakers who participate via telephone. The topic content varies from business meeting content, product presentations and demonstrations, role playing, and discussions about a prescribed topic.

Before beginning transcription, team leaders scanned each meeting session recording, identified its central topic and various other features of the meeting – for example, the number of speakers, and circulated a table with that information to the group. Transcriber team members chose to work on files with discussion topics that matched their studies or interests. For example, a team member with a finance degree chose to transcribe financial consultant sessions; another transcriber – a freelance journalist with a background in English – selected a literature discussion. This flexible, content-based approach to meetings kept LDC's team members more engaged, consistent and invested in the transcription process.

2.3 Evaluation Data

In addition to making the training data available for distribution through NIST, LDC developed a portion of the benchmark test data for this year's evaluation. The RT-07 three-hour conference room evaluation corpus includes nine excerpts from eight meeting sessions contributed by four organizations or consortia: Carnegie Mellon Institute, University of Edinburgh, National Institute of Standards and Technology, and Virginia Tech. The sessions contain an average of six participants and are twenty-two minutes long. In all cases individual head-mounted microphone (IHM) recordings were available and were used for the bulk of transcription. The meetings represent a variety of subjects, scenarios and recording conditions, but contain primarily business content.

As with the NIST Phase II corpus, team leaders scanned the audio recordings before beginning transcription and created a "meeting profile" by noting key features of each discussion: the number of speakers, discussion topic and topic-specific vocabulary, level of interaction, acoustic features, and speaker features. While assignment of the test set was random, the descriptions of the meetings provided transcribers with important information for each recording. An example of a meeting profile is shown in Table 2.

Filename	**CMU_20061115-1530**
# speakers	Four: 2 males and 2 females.
Topic of conversation	A group of transcribers discusses things that are difficult to transcribe. Some of these problems include: non-native English speakers, filled pauses, foreign languages, and proper names. They discuss potential solutions to these issues.
Vocabulary	n/a
Level of interaction	The level of interaction of this file is 2= Moderately interactive (All speakers participating, some overlap)
Acoustic features	There are minor background noises.
Speaker features	One non-native English speaker. All other speakers are native English speakers who are clearly heard and understood.
Other notes	The speakers in this file know that the file will be transcribed.

Table 2. Profile of a meeting recording in the RT-07 test set.

3 Transcription

3.1 Quick Transcription (QTR)

The goal of QTR is simply to "get the words right" as quickly as possible; to that end, the QTR methodology automates some aspects of the transcription process and eliminates most feature markup, permitting transcribers to complete a verbatim transcript in a single pass over each channel. [1] Automatic measures include pre-processing the audio signal to segment it into chunks of speech, and post-processing the transcript by running a spell check, data format check and scans for common errors. Manual audio segmentation is an integral part of careful transcription, but is very costly, accounting for 1/4 or more of the time required to produce a highly-accurate verbatim transcript. To reduce costs in QTR, LDC developed AutoSegmenter, a process that uses Entropic's ESPS library to pre-segment a speech file into speaker segments by detecting pauses in the audio stream. AutoSegmenter achieves relatively high accuracy on clean audio signals containing one speaker, and typically produces good results on the head-mounted microphone channels. When the audio is degraded in any way, the quality of automatic segmentation falls dramatically, leading to large portions of missed speech, truncated utterances, and false alarm segments – segments that may have been triggered by other participants in the room, noise, or distortion.

The QTR approach was adopted on a limited scale for English conversational telephone speech data within the DARPA EARS program [2], with real-time transcription rates of seven to ten times real-time. Team leaders monitor progress and

speed to ensure that transcripts are produced within the targeted timeframe. The resulting quick transcription quality is naturally lower than that produced by the careful transcription methodology, since accelerating the process inevitably results in missed or mis-transcribed speech; this is particularly true for difficult sections of the transcript, such as disfluent or overlapping speech sections. However, the advantage of this approach is undeniable. Annotators work ten times faster on average using this approach than they are able to work within the careful transcription methodology.

3.1.1 Quality Control

Quality assurance efforts are minimized for QTR, since the goal of this approach is to produce a transcript in as little time as possible. However, the meetings in this dataset were reviewed in a quick final pass, which involved a spell check, a data format check and contraction expansion. Transcripts were reviewed again briefly (in a one times real time pass) by a team leader for accuracy and completeness.

3.2 Careful Transcription (CTR)

For purposes of evaluating transcription technology, system output must be compared with high-quality manually-created verbatim transcripts. LDC has already defined a careful transcription (CTR) methodology to ensure a consistent approach to the creation of benchmark data. The goal of CTR is to create a reference transcript that is as good as a human can make it, capturing even subtle details of the audio signal and providing close time-alignment with the corresponding transcript. CTR involves multiple passes over the data and rigorous quality control. Some version of LDC's current CTR specification has been used to produce test data for several speech technology evaluations in the broadcast news and conversational telephone speech domains in English, Mandarin, Modern Standard and Levantine Arabic as well as other languages over the past decade. In 2004 the CTR methodology was extended to the meeting domain to support the RT-04 meeting speech evaluation. [3]

Working with a single speaker channel at a time using individual head-mounted microphone (IHM) recordings, annotators first divide the audio signal into virtual segments containing speaker utterances and noise while simultaneously labeling each speaker with a unique speaker ID. At minimum, annotators divide the audio into individual speaker turns. Turns that are longer than 10 seconds are segmented into smaller units. Speaker turns can be difficult to define in general and are particularly challenging in the meeting domain due to the frequency of overlapping speech and the prevalence of side conversations or asides that occur simultaneously with the main thread of speech. Transcribers are therefore generally instructed to place segment boundaries at natural breakpoints like breath groups and pauses, typically resulting in segments of three to eight seconds in duration.

When placing segment boundaries, transcribers listen to the entire audio file and visually inspect the waveform display, capturing every region of speech as well as isolating vocalized speaker noises such as coughs, sneezes, and laughter. Audible breaths are not captured unless they seem to convey some meaning, such as a sigh or a sharp breath. Speaker and ambient noise were annotated on separate virtual channels (VSC) in the XTrans speech annotation tool. The VSC function allows a transcriber to

attribute an undetermined number of speakers – or in this case, non-speech events for one speaker – to one audio signal. Segmenting speaker noise in this manner allowed for cleaner speech event segmentation and more accurate non-speech event information. Transcribers leave several milliseconds of silence padding around each segment boundary, and are cautious about clipping off the onset of voiceless consonants or the ends of fricatives.

After accurate segment boundaries are in place, annotators create a verbatim transcript by listening to each segment in turn. Because segments are typically around five seconds, it is usually possible to create a verbatim transcript by listening to each segment once; regions containing speaker disfluencies or other phenomena may warrant several reviews. Though no time limit is imposed for CTR, annotators are instructed to insert the "uncertain transcription" convention if they need to review a segment three or more times. A second pass checks the accuracy of the segment boundaries and transcript itself, revisits sections marked as “uncertain,” validates speaker identity, adds information about background noise conditions, and inserts special markup for mispronounced words, proper names, acronyms, partial words, disfluencies and the like. A third pass over the transcript conducted by the team leader ensures accuracy and completeness, leveraging the context of the full meeting to verify specific vocabulary, acronyms and proper nouns as required.

Transcription ends with multiple automatic and manual scans over the data to identify regions of missed speech, correct common errors, and conduct spelling and data format checks, which identify badly formatted regions of each file. These steps are described in more detail in the following section.

3.2 Quality Control

To enhance the accuracy of meeting transcription, annotators work with the separate IHM recordings of individual speakers and the merged recording of the all IHM recordings of the meeting participants. Segmentation and first-pass transcription are produced primarily from the individual IHM recordings in the manner described above. Senior annotators listen to all untranscribed regions of individual files, identifying any areas of missed speech or chopped segments using a specialized interface.

Meetings may contain highly specialized terminology and names that may be difficult for transcribers to interpret. To resolve instances of uncertainty and inconsistency, additional quality control passes are conducted using a distant or table-top microphone recording or the merged IHM recording, which conflates all individual speaker transcripts into a single session that is time-aligned with a mixed recording of all IHM channels. This merged view provides a comprehensive inspection of the consistency of terminology and names across the file, and is conducted by a senior annotator who has greater access to and knowledge of technical jargon. Senior annotators also check for common errors and standardize the spelling of proper nouns and the representation of acronyms in the transcript and across transcripts, where applicable.

The final stages of quality control involve multiple quality assurance scans, such as spell checking and syntax checking, which identifies portions of the transcript that are

poorly formatted (for example, conflicting markup of linguistic features), and expanding contractions.

4 Unique Challenges of Meeting Data

The meeting domain presents a number of unique challenges to the production of highly accurate verbatim transcripts, which motivates the application of quality control procedures as a part of the multi-pass strategy described above. One such challenge is the prevalence of overlapping speech. In meetings, overlap is extremely frequent, accounting for nearly a quarter of the speech on average.[1] Even when transcribing using a speaker's IHM recording, capturing speech in overlapping regions is difficult because other speakers are typically audible on those channels. During all stages of transcription, transcribers and team leaders devote extra attention to overlapping speech regions.

Meeting content may also present a challenge to transcribers. Much of the conference room data is collected during project discussion groups or technical meetings, and frequently involves highly-specific terminology that requires extra care and research to transcribe accurately. Furthermore, meeting attendees show very different levels of participation, and some may not speak at all during a recorded session. While this is not a major roadblock to the production of reference transcription, speakers who do not speak motivate extra care at all phases of transcription, to ensure that no speech event has been missed.

Another challenge fundamental to creating a high-quality meeting data transcript is the added volume of speech, resulting from not one or two but a half a dozen or more speakers. A typical thirty-minute telephone conversation will require twenty hours or more to transcribe carefully (30 minutes, two speakers, 20 times real-time per channel). A meeting of the same duration with six participants may require more than 60 hours producing a transcript of the same quality.

The nature of meeting speech transcription requires frequent jumping back and forth from a single speaker to a multi-speaker view of the data, which presents a challenge not only for the transcribers, but for the transcription tools they use. Many current transcription tools are not optimized for or do not permit this approach. For the most part existing transcription tools cannot incorporate output of automatic processes, and they lack correction and adjudication modes. Moreover, user interfaces are not optimized for the tasks described above.

5 Infrastructure

LDC has been using a next-generation speech annotation toolkit, XTrans, to directly support a full range of speech annotation tasks including quick and careful transcription of meetings since late 2005. XTrans, based on QT and implemented in Python and C++, utilizes the Annotation Graph Toolkit [4, 5] whose infrastructure of

[1] This is based on the RT-07 test set, where the amount of overlap ranged from 4.85%-43.04%.

libraries, applications and GUI components enables rapid development of task-specific annotation tools.

XTrans operates across languages, platforms and domains, containing customized modules for quick and careful transcription and structural spoken metadata annotation. The tool supports bi-directional text input, a critical component for languages such as Arabic. XTrans is being used for full-fledged transcription and a variety of speech annotation tasks in Arabic, Mandarin Chinese, and English at LDC.

XTrans contains user-configurable key bindings for common tasks. All commands can be issued from keyboard or mouse, depending on user preference. This user-friendly tool includes specialized quality control features; for instance speakerID verification to find misapplied speaker labels and silence checking to identify speech within untranscribed regions. The speakerID verification functions include the ability to listen to random segments – or all segments – of one speaker to identify speakerID errors and modify them as necessary. XTrans enables easy handling of overlapping speech in single-channel audio by implementing a Virtual Speaker Channel (VSC) for each speaker, not each audio channel.

To support meeting domain transcription, XTrans permits an arbitrary number of audio channels to be loaded at once. For RT-07, transcribers opened the IHM channels for each meeting recording session. They had access to distant microphone recordings when desired, and could easily toggle between the multi- and single-speaker views, turning individual channels on and off as required to customize their interaction with the data. The waveform markup display makes speaker interaction obvious, showing overlapping segments and assigning a unique color to each speaker. Figure 1 shows a transcription session that is focused on a single speaker (Subj-100).

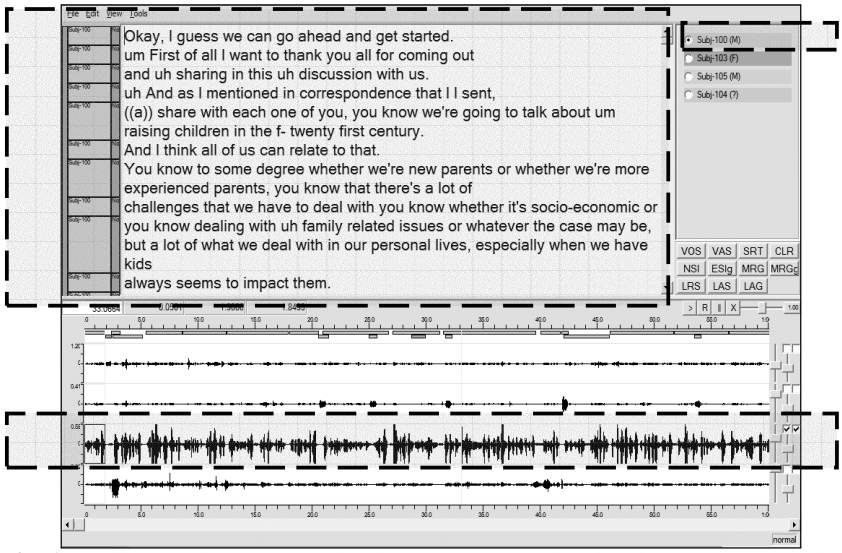

Fig. 1. Multiple audio channels, single-speaker transcript view in XTrans. Focus is on one speaker; all non-focal audio channels are turned off.

As shown in Figure 1, the transcriber may choose to show the transcript of only one speaker, and may also mute the audio recordings of the other meeting participants by de-selecting the audio output buttons to the other audio channels. When the transcriber switches to a multi-speaker view of the meeting session, the transcript for all of the meeting participants appears. The transcriber also activates the other audio output for the other speakers' recordings. The multi-speaker view is shown in Figure 2. This image also highlights the speaker featured in Figure 1.

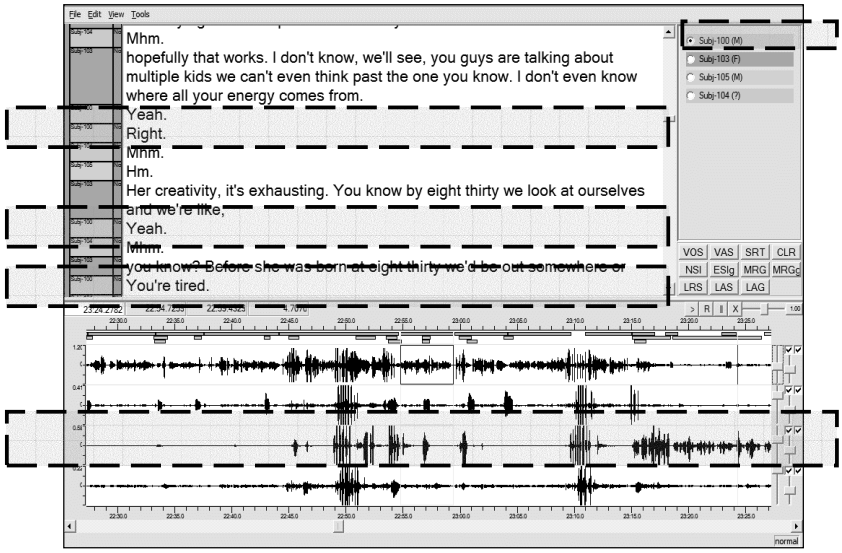

Fig. 2. Multiple audio channels, multi-speaker transcript view. Focus is on all speakers; all audio channels are activated.

As with LDC's current transcription tools, XTrans is fully integrated into LDC's existing annotation workflow system, AWS. AWS controls work (project, file) assignment; manages directories and permissions; calls up the annotation software and assigned file(s) for the user; and tracks annotation efficiency and progress. AWS allows for double-blind assignment of files for dual annotation, and incorporates adjudication and consistency scoring into the regular annotation pipeline. Supervisors can query information about progress and efficiency by user, language, data set, task, and so on.

6 Consistency pilot study

6.1 Dual transcription

For the RT-07 test set, LDC implemented the double-blind file assignment function of AWS and performed dual transcription at the first pass level, in order to understand more about inter-transcriber consistency. Seven of the nine file excerpts in the test set were dually transcribed; due to scheduling constraints, the team was not able to finish dual transcription for the entire dataset. The transcription process occurs in two distinct phases: segmentation, then transcription. To facilitate comparison between transcripts, the files were manually segmented by one transcriber. The segmentation file was copied and sent to two independent first pass transcribers. Then one of the first pass files continued through subsequent quality control passes. The file that continued through the pipeline was the one that was completed first. The workflow for the test set is shown in Figure 3.

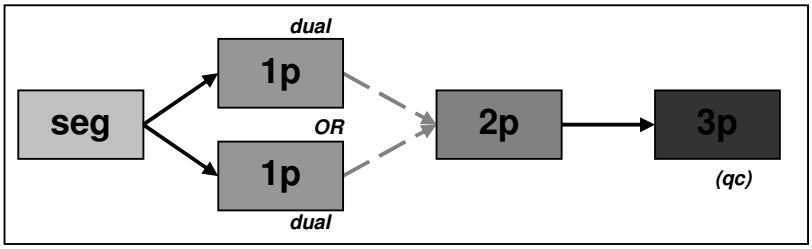

Fig. 3. RT-07 test set careful transcription workflow.

Careful first pass transcription is typically performed by junior transcribers, since the aim of the first pass is to get a verbatim transcript, ignoring markup. The transcript moves to more senior transcribers for the second pass, where markup is inserted, the transcript is carefully reviewed, proper nouns are checked, and meticulous quality control begins.

6.2 Transcript comparison

We compared the transcripts by asking transcribers to perform a form of "adjudication" by reviewing each segment of the transcripts and coding the differences. Transcribers answered a series of questions about each difference and recorded their analysis inline with the transcript. They determined whether the difference was significant or insignificant, and judged which version was correct. For significant differences, transcribers also described what caused the difference by choosing from the following options: transcriber carelessness, audio quality, the level of speaker interaction, or speaker attributes (voice quality or non-native English

speaker). Table 3 shows the key that transcribers used to analyze the differences between files.

We used a modified version of XTrans, shown in Figure 4, to view the files in parallel. The comparison view shows three versions of each line of the transcript: first pass 1, first pass 2, and first pass edit, which allows the adjudicator to correct the transcript or simply take notes about the two versions.

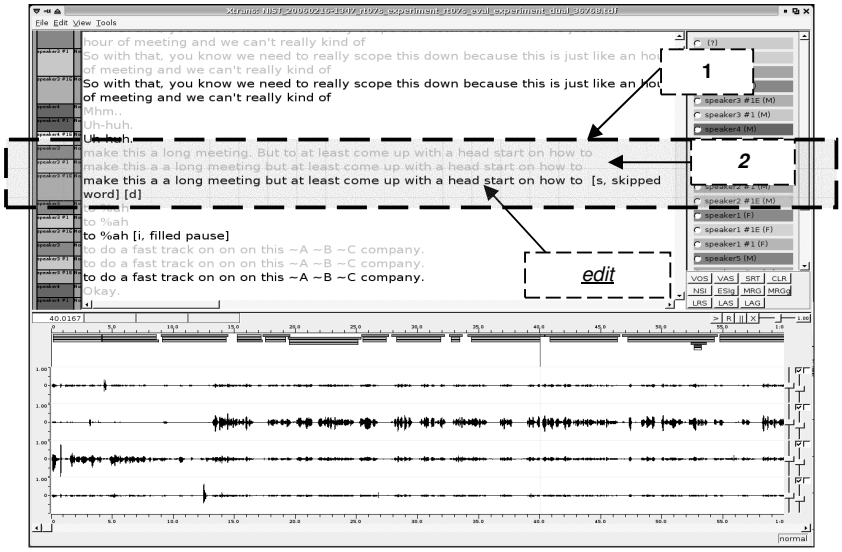

Fig. 4. Two transcripts displayed together in a customized version of the XTrans speech annotation too. A script merges the transcripts together and displays the time-aligned segment pairs together, leaving a third transcript line for comments and analysis.

6.3 Results

Though the manual comparison of transcripts was more qualitative than quantitative, we made an effort to quantify the findings of this study. To do so, we counted the number of segments, and the number of the significant and insignificant differences in each file.

Across all files, we counted a total of 3495 segments. Among those, 2392 segments differed. Of those differing segments, approximately 36.41% (871) were marked as being insignificant, which means that the differences between segments are spelling errors, punctuation and capitalization differences, lack of markup, or noise annotation. 63.59% (1521) segments that differed were marked as containing significant discrepancies. These were cases where transcriber 1 and transcriber 2 understood an utterance differently. The significant differences between transcribers range from simple – a partial word versus a full word, or “mhm” versus “uh-huh” – to

extreme – where the two transcribers wrote completely different utterances. Several examples displaying the range of comprehension deviation are included in Table 3.

analysis	*version id*	*transcript*
significant, really vs. at least, trans1	trans1	at least spell things out and possibly look them up based on that.
	trans2	really spell things out and possibly look them up based on that.
significant, three vs. Greek, trans2	trans1	We had three %um Spanish, Italian,
	trans2	We had Greek %um Spanish, Italian,
significant, missed transcription trans2	trans1	(( ))
	trans2	Yeah me too because they s-
significant, large portion left out, trans2	trans1	you know they're actually saying ^Spain and it's just you know part of sort of their phonetic make up
	trans2	you know they're actually saying ^Spain and it's just you know part of sort of their phonetic makeup to add that *schwa at the beginning and --
significant, comprehension error, trans2	trans1	Tha- that's true. ^Alex is kind of ((dead and air looking)).
	trans2	That's true, Alex is kind of debonair looking.
significant, trans2	trans1	Uh-huh.
	trans2	Mhm.
significant, misunderstood word, trans2	trans1	Yeah, and it sort of hurts because often you think this is a filled pause.
	trans2	Yeah and it sort of helps because often you think this is a filled pause
significant, filled pause, trans2	trans1	%ah yes.
	trans2	Ah yes.
significant, different segmentation and transcription, trans1	trans1	Okay. The memorial is deteriorating. I'd say the %uh problem is the memorial is deteriorating so,
	trans2	Okay.
	trans2	The memorial is deteriorating. I'd say that our problem is the memorial is deteriorating.

Table 3. Examples of discrepancies between transcribers during the consistency pilot study with RT-07 test data.

6.4 Observations

Upon closer examination of the first pass transcripts, some of the differences seem to stem from a transcriber's lack of understanding of the context of the meeting. First pass transcribers usually focus on only one speaker at a time and do not listen to all

participants at once, so these kinds of errors are understandable at this stage. Other differences are simply careless errors or comprehension errors. We did not find that one transcript in a transcript pair was always correct.

The biggest detraction to this pilot study was the segmentation. Transcripts are most easily compared when the segmentation is identical – if the segmentation differs, words are not perfectly aligned across transcripts and it becomes very difficult to see where the primary errors are. Even though the team leader instructed first pass transcribers not to modify segment boundaries, the transcript pairs *did not* end up with identical segmentation. Currently, LDC does not have mechanism for "locking" segmentation in place, which could be useful in future efforts.

We did glean a lot of positive information from this study. It proved to be an instructive management tool. Transcribers were asked to review and adjudicate a large number of careless errors, which reinforced the transcription guidelines for them. For managers, the study highlighted specific areas to underscore during training.

In the future, we would like to compare transcripts that have been transcribed in parallel from first pass through the final stages of quality control so that simple errors are resolved and only serious inconsistencies among annotators remain. We would also like to develop better tools in-house for comparing two transcripts. Analyzing each error in XTrans was constructive, but the results were not easily quantified. Researching ways to improve inter-transcriber consistency is certainly a goal in the future.

7 Transcription Rates

LDC careful transcription real-time rates for the RT-05S two-hour dataset approached 65 times real-time, meaning that one hour of data required around 65 hours of labor (excluding additional QC provided by the team leader), which is around 15 times real-time per channel, comparable with rates for BN and slightly less than that for CTS. Using XTrans to develop the RT-06S conference room data, our real-time rates dropped to under 50 times real-time per file (10 times real-time per channel). [6] Careful transcription rates for RT-07 were approximately 50 times real-time, as well.

8 Future Plans and Conclusion

LDC's planned activities include additional transcription in the meeting domain and further exploration of segmentation and annotation methods that would enhance the quality or value of reference meeting transcripts. We also plan to explore ways to make Careful Transcription more efficient. XTrans carries many built-in functions that could enrich meeting transcripts, including structural metadata and topic boundary annotation, both of which are currently being annotated under the GALE Quick-Rich Transcription (QRTR) methodology. Porting LDC's expertise in these two areas to the meeting domain may open doors to topic detection research and discourse analysis.

LDC plans to collect new data, as well. Using existing facilities at LDC developed for other research programs, meeting collection is currently opportunistic, with regularly scheduled business meetings being recorded as time allows. As new funding becomes available, we also plan to develop our collections infrastructure with additional head-mounted and lavaliere microphones, an improved microphone array, better video capability and customized software for more flexible remote recording control. While the current collection platform was designed with portability in mind, we hope to make it a fully portable system that can be easily transported to locations around campus to collect not only business meetings but also lectures, training sessions and other kinds of scenarios.

Future plans for XTrans include incorporation of video input to assist with tasks like speaker identification and speaker turn detection. We also plan to add a "correction mode" that will allow users to check manual transcripts or verify output of automatic processes including auto-segmentation, forced alignment, SpeakerID and automatic speech recognition output. Another XTrans feature which we plan to explore is the "adjudication mode", allowing users to compare, adjudicate and analyze discrepancies across multiple human or machine-generated transcripts. This would certainly provide more easily-accessible data on consistency between transcribers.

Shared resources are a critical component of human language technology development. LDC is actively engaged in ongoing efforts to provide crucial resources for improved speech technology to RT-07 program participants as well as to the larger community of language researchers, educators and technology developers. These resources are not limited to data, but also include annotations, specifications, tools and infrastructure.

Acknowledgments. We would like to thank Haejoong Lee, primary developer of the XTrans speech annotation tool, for his dedication to improving XTrans. We would also like to thank the LDC transcription team for their hard work in creating the transcripts for RT-07S.

9 References

1. Linguistic Data Consortium: RT-07 Meeting Quick Transcription Guidelines. (2007) https://projects.ldc.upenn.edu/Transcription/NISTMeet/MeetingDataQTR-V2.0.pdf.
2. Strassel, S., Cieri, C., Walker, K., Miller, D.: Shared Resources for Robust Speech-to-Text Technology, Proceedings of Eurospeech (2003).
3. Linguistic Data Consortium: RT-07 Meeting Careful Transcription Guidelines. (2007) https://projects.ldc.upenn.edu/Transcription/NISTMeet/MeetingDataCTR-V2.1.pdf.
4. Bird, S., Liberman, M.: A formal framework for linguistic annotation. Speech Communication, (2001) 33:23-60.
5. Maeda, K., Strassel, S.: Annotation Tools for Large-Scale Corpus Development: Using AGTK at the Linguistic Data Consortium. Proceedings of the 4th International Conference on Language Resources and Evaluation (2004).
6. Glenn, M., Strassel, S.: Linguistic Resources for Meeting Speech Recognition. MLMI 2005: 390-401